

## RDFS + OWL을 이용한 생물학적 데이터의 지식 표현과 추출 Knowledge Representation and Extraction of Biological Data using RDFS + OWL

이승희, 신문수, 정무영

포항공과대학교 산업공학과/제품생산기술연구소

### Abstract

Due to the lack of digitally usable standards, it has been known to be difficult to handle the biological data. For example, the name of genes and proteins changes over time or has several synonyms indicating different entities. To cope with these problems, several communities, including the Gene Ontology Consortium and PubGene are making their efforts to move science toward the semantic web vision. Although some progress has been made, its expressivity is not sufficient for full-fledged ontological modeling and reasoning. This paper suggests a methodology for representing and extracting biological knowledge by using Web Ontology Language (OWL) as an extension of Resource Description Framework Schema (RDFS). Some benefits of our approach are: (1) to ensure extended sharing of biological meta data on the Web, and (2) to enrich additional expressivity and the semantics of RDFS+OWL.

**Keywords** : Biological data, Knowledge representation/extraction, RDFS, OWL

### 1. 서론

휴먼 게놈 프로젝트(Human Genome Project)를 비롯한 다양한 생물들의 게놈 서열분석 작업이 완성되어 감에 따라 생물학적 데이터가 폭발적으로 증가하게 되었다. 이제는 유전자 생성물(gene product)인 단백질(protein)의 기능을 밝히고 생물학 프로세스를 이해하려는 학문인 프로테오믹스(Proteomics)의 시대가 도래 하고 있다. 그런데 여러 가지 생물학 데이터베이스에 존재하는 어휘나 용어가 혼재되어 쓰이거나 시간에 따라 바뀌는 동 생물학 정보를 이해하고 처리하는데 많은 어려움이 있다.

복잡한 생물학적 프로세스를 이해하기 위해서는 유전자와 유전자 생성물을 정의하고 그들 사이

의 관계를 기술하는 통일된 표준이 필요하다.

Gene Ontology Consortium(GOC)에서는 생물학 용어들 간의 관계를 정의하는 온톨로지로서 Gene Ontology(GO)를 만들어 제안하였다[1][2]. 그러나 GO는 데이터 모델이 단순한 RDF(Resource Description Framework)로 만들어져 있기 때문에 논리 기반의 온톨로지가 아니다. 따라서 복잡한 생물학적 사실들과 그들 사이의 관계를 나타내기에는 표현력이 부족하고, 생물학적 지식의 고급 검색이 불가능하다.

논리 기반의 온톨로지를 통한 지식표현은 생물학적 사실들 사이의 복잡한 상호관계를 표현하고 이해하는데 도움이 되며, 방대한 양의 생물학 데이터를 빠르고 쉽게 처리할 수 있고, 생물학적 지식의 특수성과 정확성을 증가시킬 수 있다. 또한 시맨틱(semantic) 태그가 붙은 문서들은 사실이나 사실들 사이의 관계를 추출해내는 일을 쉽게 해준다.

본 논문에서는 시맨틱 웹(semantic web)을 구성하는 기본 계층인 RDF(Resource Description Framework), RDFS(RDF Schema), 그리고 OWL(Web Ontology Language)을 간단히 설명하고, RDFS+OWL을 이용한 논리 기반의 온톨로지를 만드는 방법과 생물학적 지식을 표현하고 추출하는 방법을 제안한다.

### 2. 생물학 온톨로지

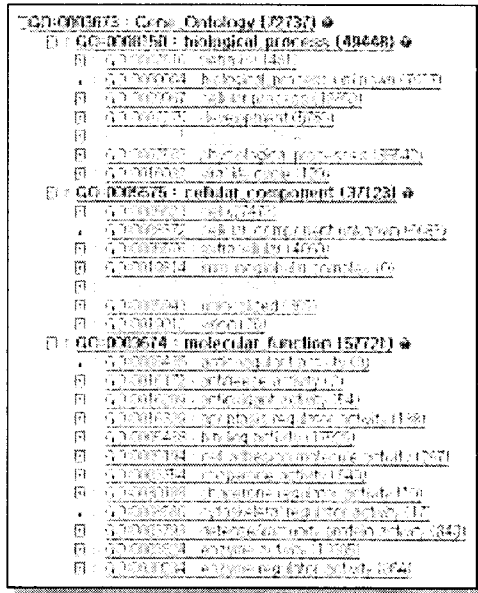
#### 2.1 온톨로지

온톨로지는 용어의 의미와 용어들 간의 관계를 명확하게 표현하기 위한 것이며, 공유된 개념화를 위한 형식적이고 분명한 명세서이다[3]. '개념화'란 세상에 존재하는 임의의 현상에 대한 추상화 모델을 통해 그 현상과 관련된 개념들을 확인해 준다는 것을 의미한다. 또한 '분명한'의 의미는 개념들의 타입과 그 개념들을 사용하는 데 있어서의 제약을 분명하게 정의해야 함을 의미하며, '형식적인'의 의미는 일련의 프로그램으로 작동하는 컴퓨터가 정보를 모호함 없이 통역하고 처리할 수 있어야 한다는 것을 의미한다. '공유된'의 의미는 온톨로지가 일종의 합의된 지식으로서 어떤 개인에 제한된 것이 아니라 단체에 의해 받아들여져야 한다는 뜻이다. 즉 온톨로지는 임의의 현상에 대한 추상화 모델을 통해 컴퓨터가 모호함 없이 정보를 처리할 수 있도록 하

는 보편적 지식이다.

## 2.2 Gene Ontology (GO)

GOC에서 제안한 생물학 온톨로지인 GO는 크게 분자기능, 생물학적 프로세스, 세포요소 세 가지 온톨로지로 구성되어 있다. 각각의 온톨로지는 다시 하위 수준의 세부 온톨로지들로 나누어지는 수직 계층구조를 가진다. [그림 1]은 GO의 구조를 나타내고 있으며, 괄호 안의 숫자는 각 온톨로지에서의 정의하고 있는 용어의 개수이다.



[그림 1] Gene Ontology(GO)의 구조

GO에서는 용어들 사이의 구성 요소 관계와 인스턴스 관계를 각각 "part of"와 "isa"로 나타낸다. 이러한 구성 요소 관계를 통해 복잡한 생물학적 지식의 표현이 가능하다. 예를 들어 감수분열시의 염색체 위치처럼 요소(component)가 여러 장소에 존재할 수 있음을 나타내어야 하는 경우를 고려할 수 있다. 효모의 경우 감수 분열시 염색체가 핵에 존재하는데 비해서 파리나 쥐의 경우에는 염색체가 세포질에 존재한다. 또한 많은 생물들의 경우 염색체가 핵에 위치해 있지 않다. 따라서 일반적인 생물의 염색체는 [그림 2]에서와 같이 세포질과 핵에 존재하는 염색체를 간기(interphase)와 유사분열(mitotic)시에 존재하는 염색체의 조합으로 나타내어야 한다. [그림 2]에서 실선은 "part of", 점선은 "isa", 쇄선은 경로(pathway)를 나타낸다. 그러나 이러한 수준의 관계 표현만으로는 표현력의 한계로 인해 고급 지식의 표현이 불가능하다. GO의 한계점을 구성 요소 온톨로지 별로 기술하면 다음과 같다.

### 분자기능 온톨로지

분자 기능이란 RNA와 단백질과 같은 유전자 생성물의 기능 또는 활동을 의미한다. GO의 분자기능 온톨로지는 단어의 끝에 'activity'를 붙여 용어를

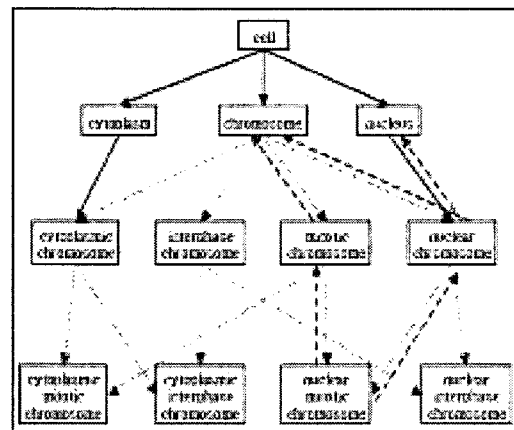
표현한다. GO에서는 유전자 생성물 중에서 그 자체가 기능이 되는 경우와 기능이 되지 않는 경우를 구별해야 한다. 예를 들어 리보솜 그 자체는 기능이 될 수 없지만, 리보솜의 구조적 요소인 단백질은 기능으로 간주한다. 그러나 이러한 구별은 의미론적인 혼란을 가져올 수도 있다. 예를 들면, 단백질 복합체인 일반전사인자(general transcription factor)는 [그림 3]이 나타내는 것처럼 TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIF, TFIIFH 등의 단백질로 이루어져 있다. 이 중에서 TFIID는 TAFs와 TBP의 복합체이다. TFIID의 경우처럼 스스로 복합체인 동시에 다른 복합체의 구성 요소인 용어들은 분자 기능과 세포 요소에 둘 다 해당되게 된다. 또한, 일반적으로 유전자 생성물은 그것의 분자 기능을 따서 이름을 짓기 때문에 유전자 생성물과 그것의 분자 기능을 구별하는 것은 쉽지 않다(예, 효소).

### 생물학적 프로세스 온톨로지

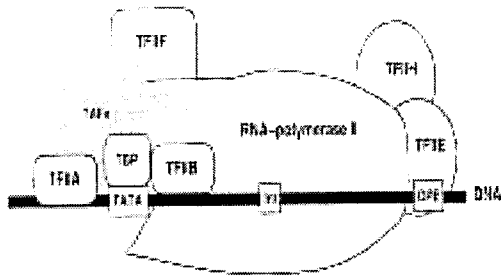
생물학적 프로세스는 어떤 목적을 달성하기 위해 여러 기능을 필요로 한다. GO의 생물학적 프로세스는 한 개 이상의 분자 기능이 참여하여 특정한 목적을 수행하는 과정을 의미한다. 그러나 GO의 생물학적 프로세스는 생물학적 경로의 의미는 포함하지 않고, 분자 기능들과 생물학적 프로세스들 사이에 분명하고 정확한 관계를 나타내지 못하고 다만 한 프로세스에 포함된 분자 기능들은 어떤 것들이 있는지만을 기술할 수 있을 뿐이다.

### 세포요소 온톨로지

세포요소 온톨로지는 개개의 단백질이나 핵산이 아닌, 여러 서브유닛을 가진 효소(multi subunit enzyme)나 단백질 복합체(protein complex)를 표현하며(예, 리보솜 또는 골지체), 세포내에서 유전자 생성물이 발견되는 위치(예, 핵막)를 나타낸다. 세포 요소에 관련된 용어들은 세포 구조에 대한 지식을 반영하고 있다고 볼 수 있다. GO에서는 유전자 생성물에서와 마찬가지로 세포 요소 중에서도 그 자체가 기능이 되는 것과 그렇지 않은 것을 구별해야 한다(예, hedgehog가 되는 것 자체는 기능이라고 할 수 없지만 hedgehog의 수용체가 되는 것은 기능이라고 할 수 있다).



[그림 2] 복잡한 생물학적 지식의 예



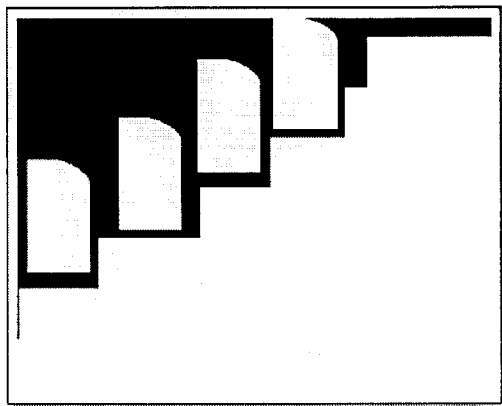
[그림 3] 일반전사인자(general transcription factor)

GO는 용어의 이름의 모호한 경우(예, 유전자 생성물과 분자 기능의 이름이 비슷한 경우) 이를 구분할 수 있는 해결책이 없으며, 동의어 처리가 불가능하다. 또한 관계를 기술하는 어휘가 부족하기 때문에 용어들 사이의 분명하고 정확한 관계 표현이 어렵고, 돌연변이가 개체에서만 일어나는 프로세스를 표현하지 못하며, 같은 생성물로 가는 서로 다른 경로/프로세스들은 하나의 경로의 인스턴스로 취급함으로써 문제를 쉽게 단순화시켜버리는 단점이 있다.

### 3. RDFS+OWL을 이용한 생물학 지식 표현

#### 3.1. RDF, RDFS, OWL

시맨틱 웹은 웹이 제공하고 있는 정보를 잘 정의된 온톨로지를 기반으로 하여 표현함으로써, 물리적으로 논리적으로 분산되어 있는 애플리케이션 간의 상호 운용성을 제공하기 위한 것이다. 즉 시맨틱 웹에서는 서로 다른 데이터 구조를 갖고 있는 애플리케이션들도 온톨로지를 통해 상대방의 정보를 이해하고 처리할 수 있다. Tim Berners-Lee는 시맨틱 웹을 '잘 정의된 의미를 가진 정보를 제공하여 기계와 인간의 협력을 더욱 향상시키는 웹의 확장'이라고 정의하고 있다[4]. [그림 4]는 Tim Berners-Lee가 제안한 시맨틱 웹 구현을 위한 아키텍처이다.



[그림 4] 시맨틱 웹 아키텍처[4]

W3C에서는 시맨틱 웹의 구현을 위해 웹 온톨로지 언어로서 OWL(Web Ontology Language)을 제안하

고 있다[5]. OWL은 RDF(Resource Description Framework), RDFS (RDF Schema)와는 다르게 논리관계의 표현능력을 제공한다. RDF와 RDFS는 각각 기본적 데이터 모델과 데이터 모델링 구조를 정의하며, 사실 관계의 표현만 가능할 뿐이다. 따라서 추론을 통한 고급정보의 표현을 위해서는 OWL의 사용이 필요하다. RDF, RDFS 그리고 OWL의 주요 특징을 기술하면 다음과 같다.

#### RDF(Resource Description Framework)

RDF는 시맨틱 웹의 구현을 위한 최하위 계층의 역할을 담당하며, 기계가 처리할 수 있는 간단한 데이터 모델과 컴퓨터가 처리할 수 있는 표준화된 신택스(syntax)를 제공한다. RDF의 기본적인 데이터 모델은 주어(subject), 속성(property) 그리고 목적어(object), 세 가지 객체 타입으로 구성된다. 주어는 자원(resource)을 의미하며, 웹 페이지의 전체 혹은 부분, 웹 페이지들의 집합 그리고 인쇄된 책자처럼 웹을 통해 직접 접근할 수 없는 객체를 URI로 나타낸다. 또한 속성은 주어, 즉 자원을 기술하기 위한 특정한 측면, 성질, 특징, 관계를 나타내며, 목적어는 속성값을 나타내거나 속성에 해당하는 자원을 나타낸다.

#### RDFS(RDF Schema)

RDFS는 데이터 모델링의 구조로서 클래스와 속성을 정의하고 그들 사이의 관계를 정의한다. RDFS가 정의하는 클래스는 크게 다음 세 가지로 볼 수 있다; 1) rdfs:Resource, 2) rdf:Property, 3) rdfs:Class. RDFS로 표현되는 모든 것은 rdfs:Resource는 RDFS로 표현할 수 있는 모든 것을 인스턴스로 갖는 클래스이며, rdf:Property는 rdfs:Resource 클래스의 인스턴스의 특징을 나타내기 위해 사용되는 모든 속성들의 클래스이다. 마지막으로 rdfs:Class는 RDFS에서 개념을 나타내기 위한 클래스이다. 즉, 모든 개념들은 rdfs:Class의 인스턴스이다.

RDFS가 정의하는 속성은 크게 다음 세 가지로 나눌 수 있다; 1) rdf:type, 2) rdfs:subClassOf, 3) rdfs:subPropertyOf. rdf:type는 resource와 클래스 사이의 instance-of 관계를 나타낸다. 하나의 resource는 하나 이상의 클래스의 인스턴스이다. rdfs:subClassOf는 클래스들 사이의 포함 계층관계를 나타낸다. 한 클래스는 여러 클래스의 서브클래스가 될 수 있다. 그러나 클래스는 그 자신(혹은 자신의 서브클래스)의 서브클래스가 될 수는 없다. 즉, 상속그래프에서 순환 관계가 없다. rdfs:subPropertyOf는 속성들 사이의 포함 계층관계를 나타낸다. 속성 P<sub>2</sub>가 P<sub>1</sub>의 rdfs:subPropertyOf이고 R이 속성 P<sub>2</sub>의 값으로 V를 갖는다고 하면 R은 속성값이 V인 P<sub>1</sub>도 속성으로 가진다. 역시 상속그래프에서 순환 관계가 없다.

또한 RDFS는 클래스와 속성들 사이의 관계를 정의하기 위해서 다음과 같은 제약 조건들을 정의한다. 1) rdfs:ConstraintResource, 2) rdfs:ConstraintProperty, 3) rdfs:range, 4) rdfs:domain. rdfs:ConstraintResource는 모든 제약 조건을 인스턴스로 갖는 클래스이며, rdfs:ConstraintProperty는 rdfs:ConstraintResource의 부분집합이다. rdfs:ConstraintResource와 rdf:Property는 제약을 정의하기 위한 모든 속성을 포함한다. rdfs:range, rdfs:domain은 rdfs:ConstraintProperty의 인스턴스로서 속성의

range와 domain을 제한하기 위해 사용된다. 한 속성에 대해서 두개 이상의 range 제약을 들 수 없다. domain의 경우는 domain들의 합집합이 허용된다.

**OWL(Web Ontology Language)**

RDF, RDFS는 형식적인 시맨틱을 위한 규정을 포함하지 않고 있고, 추론을 하기에는 표현력이 불충분하다. OWL은 RDFS보다 시맨틱을 나타내는 표현력이 뛰어나며 클래스와 속성에 다양한 논리를 제공한다. OWL에서 모든 객체는 owl:Thing 클래스의 인스턴스이고 사용자가 정의한 모든 클래스는 owl:Thing 클래스의 서브클래스가 된다. OWL에서는 속성에 대한 다양한 특성을 기술하는 어휘들을 제공한다. 예를 들어, 속성이 가질 수 있는 값의 범위를 제한하는 어휘로는 owl:allValuesFrom, owl:someValuesFrom, owl:hasValue 등이 있고, 속성이 가질 수 있는 값의 개수를 제한하는 어휘로는 owl:maxCardinality, owl:minCardinality, owl:cardinality 등이 있다. 또한 속성간의 논리적 관계를 표현하는 어휘로는 owl:equivalentProperty, owl:inverseOf, owl:TransitiveProperty, owl:SymmetricProperty 등이 있다. 클래스의 논리적 관계를 표현하는 어휘로는 owl:intersectionOf, owl:unionOf, owl:complementOf 등이 있다.

OWL에 있는 속성 제한, 속성간의 논리적 관계, 클래스 간의 논리적 관계를 사용하여 생물학 온톨로지의 스키마를 만들면 복잡한 생물학적 사실들 간의 관계를 표현하는데 유용하다.

**3.2 RDFS+OWL을 이용한 생물학 지식표현**

XML은 사용자 마음대로 태그를 정의할 수 있고 구조화된 문서를 위한 신택스를 제공하지만 문서에 시맨틱이 없다. RDF는 객체와 객체들 간의 관계를 위한 데이터 모델로서 간단한 시맨틱은 제공하지만 클래스와 속성의 논리적인 특성을 표현하는 어휘가 부족하다. OWL에서는 속성, 클래스 간의 관계와 속성의 타입, 특정 등을 기술하는 다양한 어휘를 제공함으로써 시맨틱의 표현력이 풍부해지고 객체간의 논리적 관계를 보다 잘 나타낼 수 있다. 온톨로지는 용어의 의미와 용어들 간의 관계를 명확하게 표현할 수 있어야 하므로 OWL의 속성 제한, 속성간의 논리적 관계, 클래스 간의 논리적 관계를 사용한 온톨로지는 보다 다양한 생물학적 사실들 간의 관계를 표현할 수 있게 된다.

본 논문에서는 OWL의 속성 제한, 속성간의 논리적 관계, 클래스 간의 논리적 관계를 사용하여 복잡한 생물학적 사실들 간의 관계를 표현하고자 한다. 특히 RDFS+OWL을 이용한 지식표현의 예로써 서브 클래스간의 disjoint한 관계를 표현하는 경우와 속성의 특성을 제한하는 경우를 들어 설명하고자 한다.

**서브 클래스간의 Disjoint 관계표현**

생물의 유전자는 원핵생물(prokaryote)의 유전자와 진핵생물(eukaryote)의 유전자로 나눌 수 있다. 유전자는 단백질로 번역이 되는 부분(coding region)과 번역이 되지 않는 부분(non-coding region)으로 되어 있다. 원핵생물의 경우 유전자의 구조가 매우 단순해서 유전자인 부분은 모두 단백질로 번역이 되고, 유전자가 아닌 부분은 단백질로 번역이 되지 않는다. 진핵생물의 유전자는 원핵생물의 유전자에 비해 구조가 복잡하다. 진핵생물에서

유전자가 아닌 부분은 원핵생물과 마찬가지로 단백질로 번역이 되지 않지만 유전자인 부분은 단백질로 번역이 되는 부분인 엑손(exon)과 단백질로 번역이 되지 않는 부분인 인트론(intron)으로 구성되어 있다. 따라서 [그림 5]에서와 같이 원핵생물의 유전자와 진핵생물의 유전자 사이의 관계는 owl:disjointWith로 나타내고 엑손과 인트론은 rdf:parseType="Collection"을 사용하여 진핵생물의 유전자를 구성하는 부분임을 나타낼 수 있다.

**속성특성의 제한 표현**

순서가 중요하지 않은 집합을 표현하는 경우를 예로써 고려할 수 있다. TFIIA, TFIIIB, TFIIID, TFIIIE, TFIIIF, TFIIIH, 등 여러 단백질들의 복합체인 일반적인 전사 인자(general transcription factor)는 DNA에서 RNA로 전사가 일어날 때 RNA 중합효소 II와 함께 기본 전사 도구(basal transcription apparatus)를 구성하여 유전자의 promoter 서열을 인식해서 RNA 중합효소가 유전자의 앞부분에 결합하여 전사를 시작하도록 도와준다. [그림 6]은 여러 단백질의 복합체인 일반 전사 인자를 OWL의 Collection을 사용해서 표현한 예이다.

```
<owl:Class rdf:ID="ProkaryoticGene">
  <rdf:type rdfs:ClassOf rdfs:resource="#Gene"/>
  <owl:disjointWith rdfs:resource="#EukaryoticGene"/>
</owl:Class>

<owl:Class rdf:ID="EukaryoticGene">
  <rdf:type rdfs:ClassOf rdfs:resource="#Gene"/>
  <rdf:type owl:CollectionOf rdfs:resource="#Collection">
    <rdf:type owl:ClassOf rdfs:resource="#Exon"/>
    <rdf:type owl:ClassOf rdfs:resource="#Intron"/>
  </rdf:type>
</owl:Class>
```

[그림 5] RDFS+OWL을 이용한 지식표현의 예(1)

```
<owl:ObjectProperty rdf:ID="isAComponentOf">
  <rdf:type owl:ObjectPropertyOf rdfs:resource="#GeneProkaryotic"/>
  <rdf:type owl:ObjectPropertyOf rdfs:resource="#CellularComponent"/>
</owl:ObjectProperty>

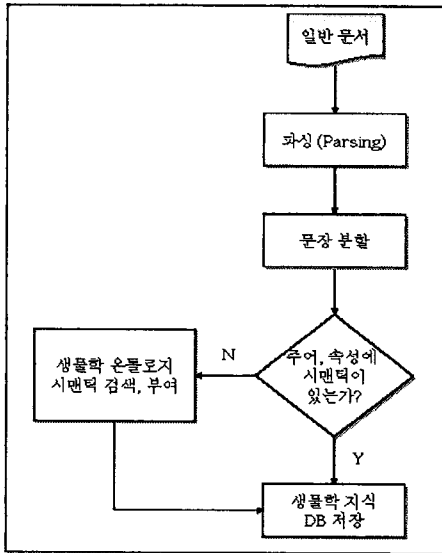
<owl:Class rdf:ID="GeneralTranscriptionFactor">
  <rdf:type owl:ClassOf rdfs:resource="#ProteinComplex"/>
  <owl:disjointWith rdfs:resource="#Collection">
    <owl:item rdf:resource="#TFIIA"/>
    <owl:item rdf:resource="#TFIIIB"/>
    <owl:item rdf:resource="#TFIID"/>
    <owl:item rdf:resource="#TFIIIE"/>
    <owl:item rdf:resource="#TFIIIF"/>
    <owl:item rdf:resource="#TFIIH"/>
  </owl:disjointWith>
</owl:Class>

<owl:Class rdf:ID="TFIID">
  <rdf:type owl:ClassOf rdfs:resource="#Component"/>
  <owl:subclassOf rdfs:resource="#Collection">
    <owl:item rdf:resource="#TAFs"/>
    <owl:item rdf:resource="#TBP"/>
  </owl:subclassOf>
</owl:Class>
```

[그림 6] RDFS+OWL을 이용한 지식표현의 예(2)

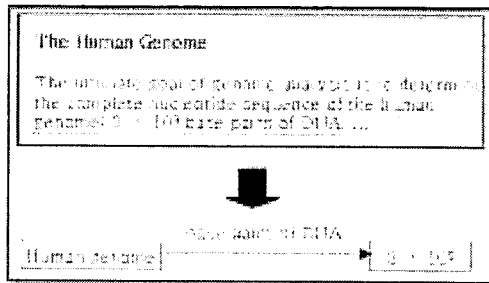
### 3.3 생물학적 데이터의 지식추출

[그림 7]은 생물학적 정보를 담고 있는 텍스트 형식의 일반 문서에서 생물학적 지식을 추출하는 방법을 도식화하여 나타낸 것이다. 우선 문서의 정보를 단어 단위로 파싱(parsing)해서 파싱된 단어가 시맨틱을 가졌을 경우 주어(subject)인지, 속성(property)인지, 목적어(object)인지 구분해서 RDF 구문에 넣고 시맨틱을 가지지 않을 경우 생물학 온톨로지에서 그 단어를 검색한다. 생물학 온톨로지에 그 단어가 존재하면 주어, 속성, 목적어를 체크하는 과정을 반복하고, 존재하지 않으면 GO에서 그 단어에 시맨틱을 부여할 것인지를 결정한다. 시맨틱을 부여하기로 했으면 생물학 온톨로지에 그 단어를 추가하고 아니면 그 다음 단어를 파싱한다.

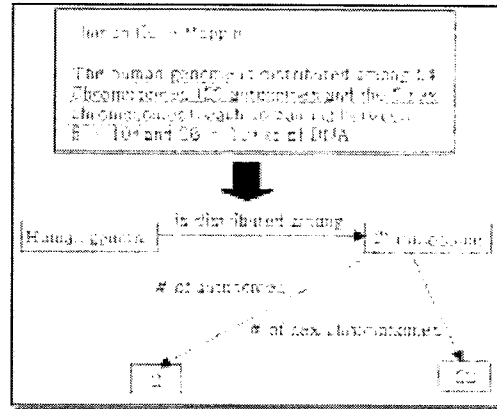


[그림 7] 생물학적 지식 추출 순서도

[그림 8]과 [그림 9]는 생물학적 정보를 담고 있는 일반 문서에서 직접 지식을 추출하는 예를 보여주고 있다. 이 예에서 일반 문서는 인간 게놈이  $3 \times 10^9$  개의 염기쌍으로 이루어진 DNA라는 생물학적 지식을 담고 있다. [그림 10]은 [그림 8]과 [그림 9]에서 얻은 지식을 표현한 예이다.



[그림 8] 일반 문서에서 생물학 지식 추출의 예(1)



[그림 9] 일반 문서에서 생물학 지식 추출의 예(2)

```

<owl:ObjectProperty rdf:ID="IsDistributedAmong">
  <rdf:domain rdf:resource="#Genome"/>
  <rdf:range rdf:resource="#Chromosome"/>
</owl:ObjectProperty>

<owl:Class rdf:ID="HumanGenome">
  <rdf:subClassOf rdf:resource="http://www.example.org/biology/#Genome"/>
  <rdf:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasBasePairs"/>
      <owl:cardinality>3*10^9</owl:cardinality>
    </owl:Restriction>
  </rdf:subClassOf>
  <rdf:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasAutosomes"/>
      <owl:cardinality>22</owl:cardinality>
    </owl:Restriction>
  </rdf:subClassOf>
  <rdf:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasSexChromosome"/>
      <owl:cardinality>2</owl:cardinality>
    </owl:Restriction>
  </rdf:subClassOf>
</owl:Class>
  
```

[그림 10] [그림 8], [그림 9]에서 추출된 지식의 표현

### 4. 결론 및 추후 연구과제

이번 연구에서는 RDFS+OWL을 이용해서 복잡한 생물학적 사실들 간의 상호관계 또는 지식을 표현하는 방법과 생물학적 정보를 담고 있는 일반 문서에서 직접 지식을 추출하는 방법을 제시하였다.

본 연구에서 제안한 RDFS+OWL을 이용한 지식표현 및 추출방법은 생물학적 사실들 사이의 다양한 논리관계의 표현능력을 제공함으로써 생물학적 지식의 특수성과 정확성을 증가시키고, 복잡한 생물학 프로세스를 이해하는데 도움을 주고, 추론을 통한 고급정보의 표현을 가능하게 해줄 것이다.

추후에는 일반 문서에서 지식을 추출할 때 여러 단어로 이어진 단어를 처리하는 문제를 해결하는 방안을 검토하고, 여러 가지 고급 검색을 가능하게 하기 위한 스키마 구성에 대해서 연구가 필요하다.

한국경영과학회/대한산업공학회 2003 춘계공동학술대회  
2003년 5월 16일-17일 한동대학교(포항)

#### 감사의 글

본 논문은 한국과학재단의 목적기초연구과제 (2001-1-31500-005-1) 와 교육인적자원부의 BK21 과제 지원으로 수행되었습니다. 이에 감사드립니다.

#### References

- [1]<http://www.geneontology.org/>
- [2]Gene Ontology Consortium, "Creating the Gene Ontology Resource: Design and Implementation", Genome Research, pp. 1425-1433, 2001.
- [3]<http://www.geneontology.org/doc/GO.usage.html>
- [4]Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001.
- [5]<http://www.w3.org/>